

## Errors in reported degrees and respondent driven sampling: Implications for bias<sup>☆</sup>



Harriet L. Mills<sup>a,\*</sup>, Samuel Johnson<sup>b</sup>, Matthew Hickman<sup>c</sup>, Nick S. Jones<sup>b</sup>, Caroline Colijn<sup>b</sup>

<sup>a</sup> MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Dynamics, Imperial College London, St Mary's Hospital, Norfolk Place, London W2 1PG, UK

<sup>b</sup> Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

<sup>c</sup> School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK

### ARTICLE INFO

#### Article history:

Received 11 April 2014

Received in revised form 5 June 2014

Accepted 5 June 2014

Available online 21 June 2014

#### Keywords:

Respondent driven sampling

At-risk populations

Contact network size

### ABSTRACT

**Background:** Respondent Driven Sampling (RDS) is a network or chain sampling method designed to access individuals from hard-to-reach populations such as people who inject drugs (PWID). RDS surveys are used to monitor behaviour and infection occurrence over time; these estimations require adjusting to account for over-sampling of individuals with many contacts. Adjustment is done based on individuals' reported total number of contacts, assuming these are correct.

**Methods:** Data on the number of contacts (degrees) of individuals sampled in two RDS surveys in Bristol, UK, show larger numbers of individuals reporting numbers of contacts in multiples of 5 and 10 than would be expected at random. To mimic these patterns we generate contact networks and explore different methods of mis-reporting degrees. We simulate RDS surveys and explore the sensitivity of adjusted estimates to these different methods.

**Results:** We find that inaccurate reporting of degrees can cause large and variable bias in estimates of prevalence or incidence. Our simulations imply that paired RDS surveys could over- or under-estimate any change in prevalence by as much as 25%. These are particularly sensitive to inaccuracies in the degree estimates of individuals with who have low degree.

**Conclusions:** There is a substantial risk of bias in estimates from RDS if degrees are not correctly reported. This is particularly important when analysing consecutive RDS samples to assess trends in population prevalence and behaviour. RDS questionnaires should be refined to obtain high resolution degree information, particularly from low-degree individuals. Additionally, larger sample sizes can reduce uncertainty in estimates.

© 2014 The Authors. Published by Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Respondent Driven Sampling (RDS) is a network or chain sampling method designed to access populations of individuals that are “hard-to-reach.” For example, people who inject drugs (PWID) or commercial sex workers (CSW) are “hidden populations,” without a recognised sampling frame and often unwilling to be identified. RDS is commonly used to deliver health education as well as to sample these populations to understand the spread of disease, the community's behavioural patterns, use of interventions, and individuals' responses to risk (Abdul-Quader et al., 2006; Broadhead

et al., 2002, 1998; Des Jarlais et al., 2007; Johnston et al., 2008; Malekinejad et al., 2008; Robinson et al., 2006). RDS works as follows: a number of individuals (*seeds*) are recruited at random from the population. (We note that in reality, seeds are preferentially selected to optimise recruitment and to increase the diversity in the sample.) These individuals are interviewed and given a set number of tokens to recruit their contacts. Successfully recruited contacts are interviewed and given tokens to recruit the next wave of individuals. The process continues until either recruitment fails or the target number of recruits is reached. RDS carries the significant advantage that no-one is asked to name contacts directly; participants are invited through their contacts and can choose whether to participate. As such, it is the current method of choice for accessing hard-to-reach populations, not only to deliver public health interventions but to gather data to estimate the prevalence and incidence of infections such as HCV and increasingly HIV (for example, Hope et al., 2010; Iguchi et al., 2009; Sypsa et al., 2014).

<sup>☆</sup> Supplementary material can be found by accessing the online version of this paper. Please see Appendix A for more information.

\* Corresponding author. Tel.: +44 20 7594 1379.

E-mail address: [harriet.l.mills@gmail.com](mailto:harriet.l.mills@gmail.com) (H.L. Mills).

Accordingly, understanding sources of variability and bias in RDS estimates is increasingly important.

Inevitably, individuals with a high number of contacts will be over-sampled in RDS studies, as these individuals know more people in the target population and therefore are more likely to be recruited. (For those who may doubt the severity of this over-sampling, it can be demonstrated in simulations with minimal assumptions, and is more severe in networks with greater variability in the numbers of contacts; see Supplementary Text S1 and Fig. S1.) In addition, as individuals with high numbers of contacts may be at greater risk of becoming infected (through contact with a larger network of injectors) and also may have a greater infecting risk (e.g., being homeless; Friedman et al., 2000), the prevalence in the sample is expected to be higher than the prevalence in the at-risk community. It is therefore necessary to adjust for this bias when estimating an infection's prevalence or incidence using RDS data (Gile and Handcock, 2010; Goel and Salganik, 2010; Heckathorn, 2007; Salganik and Heckathorn, 2004; Volz and Heckathorn, 2008). The estimate  $\hat{\mu}$  is [40]:

$$\hat{\mu} = \frac{\sum_{i=1}^n (f_i/d_i)}{\sum_{i=1}^n (1/d_i)} \quad (1)$$

where  $n$  is the sample size,  $f_i$  is the trait (e.g.,  $f_i = 1$  if the individual is infected and 0 if not) and  $d_i$  is the estimated number of contacts, or degree, of individual  $i$  (see Supplementary Text S2). Naturally, if infection were not correlated with degree, then this adjustment would not have any effect on the estimate.

An individual's degree is generally their own estimate of the number of other individuals they know by name that they have seen in a set time period, who also belong to the population being sampled (e.g., who are also PWID or CSW or other target population). This number is therefore an estimate of the number of individuals they may recruit, and also of the number of contacts relevant for the transmission of disease. However, degree may be difficult to estimate accurately as well as being dynamic in time (Brewer, 2000; Rudolph et al., 2013). Individuals may only roughly know their degree, may only recall or count close contacts or may intentionally give an inaccurate estimate, for example to hide how at risk they are or to boost their apparent popularity (desirability bias; Fisher, 1993). Degree bias or digital preference is particularly relevant in the reporting of sexual or drug use behaviours, where individuals may be uncertain or wish to avoid association with illegal or undesirable activities (Fenton et al., 2001; Schroder et al., 2003). One of the assumptions underpinning RDS and the adjustment methods is that respondents accurately report their degree. As noted by several authors, inaccuracy in degree constitutes a source of sampling bias in the adjustment procedure (Goel and Salganik, 2009; Johnston et al., 2008; Rudolph et al., 2013; Salganik and Heckathorn, 2004; Wejnert, 2009), yet to the best of our knowledge there has been no study examining the extent to which this might be important in the interpretation of RDS surveys.

There have been several other concerns about the extent to which real RDS studies match the idealised assumptions underlying the statistical estimators. Heckathorn showed that under ideal conditions, RDS samples are Markov chains whose stationary distribution is independent of the choice of seeds (Heckathorn, 1997, 2002; Salganik and Heckathorn, 2004). However, there have been concerns that preferential referral behaviour of respondents (Bengtsson and Thorson, 2010), short recruitment chains compared to the length needed for the Markov chain to reach equilibrium, and the difference between with-replacement random walk models and without-replacement real-world samples could lead to bias in RDS estimates (Gile and Handcock, 2010).

Here, we explore how reported degree data might arise from a true underlying distribution due to individuals rounding their

numbers of contacts up or down to multiples of 5, 10 and 100. We use simulations of RDS to investigate the potential bias caused by inaccurately reporting degrees and compare it to other issues researchers have raised about RDS (including the difference between with- and without- replacement sampling, multiple seed individuals and multiple recruits per individual).

## 2. Methods

### 2.1. Data

We base our methodological work on two cross-sectional RDS studies of PWID in Bristol, UK, in 2006 ( $n = 299$ ) and 2009 ( $n = 292$ ), described elsewhere (Hickman et al., 2009; Hope et al., 2011, 2013; Mills et al., 2012). They used the same questionnaire and recruited individuals who injected in the last 4 weeks. The results were used to estimate trends in HCV prevalence and incidence in this population. We analyse the reported contact numbers (degrees) from both surveys.

### 2.2. RDS simulations

We generate contact networks of individuals with a defined degree distribution using the configuration model (Newman, 2003). The contact number distribution in the Bristol data is approximately long-tailed in that reported numbers vary by several orders of magnitude, so we used a long-tailed degree distribution (power law with an exponential cut off, mean degree of 10) in the simulations. We simulate the transmission of a pathogen (SIS) across the network and after a set time we simulate an RDS survey. Details of the network and transmission model are in the Supplementary Text For comparison we present results for a network with a Poisson degree distribution, where there is much less variation in degrees (Supplementary Text S3).

We determine the impact of inaccurate degrees on the prevalence estimate by re-computing the estimate in Eq. (1) using  $d_i = \hat{d}_i + \Delta d_i$ , where  $\hat{d}_i$  are the individuals' correct degrees in the network, and  $\Delta d_i$  correspond to inaccuracies in these degrees. We consider five different rounding schemes to mimic patterns seen in data: (1) round all degrees up to the nearest 5, (2) round all degrees up to the nearest 10, (3) increase every degree by 5, and finally two methods to directly mimic patterns seen in the Bristol data (Fig. 1). These are (4) round all degrees between 10 and 100 to the nearest 10, and degrees greater than 100 to the nearest 100; and (5) similar, but individuals with degrees less than 10 are given a different degree between 1 and 10, chosen according to the distribution seen in the Bristol data.

We simulate a number of variations of RDS. First, we take a *standard "real world"* RDS sample: individuals recruit a number of their contacts to the sample, where this number is chosen from a Poisson distribution, mean 1.5 and limited to between [0,3] (and cannot be larger than their total number of contacts). Individuals cannot be sampled more than once. We compare this to idealised RDS, or Markov process RDS: there are *multiple seeds*, seeds recruit one individual only at random from their contacts and sampling is with replacement. We also use variants of this method, allowing *multiple tokens* (recruits), and *without replacement*. In all of our variants, seeds are chosen at random.

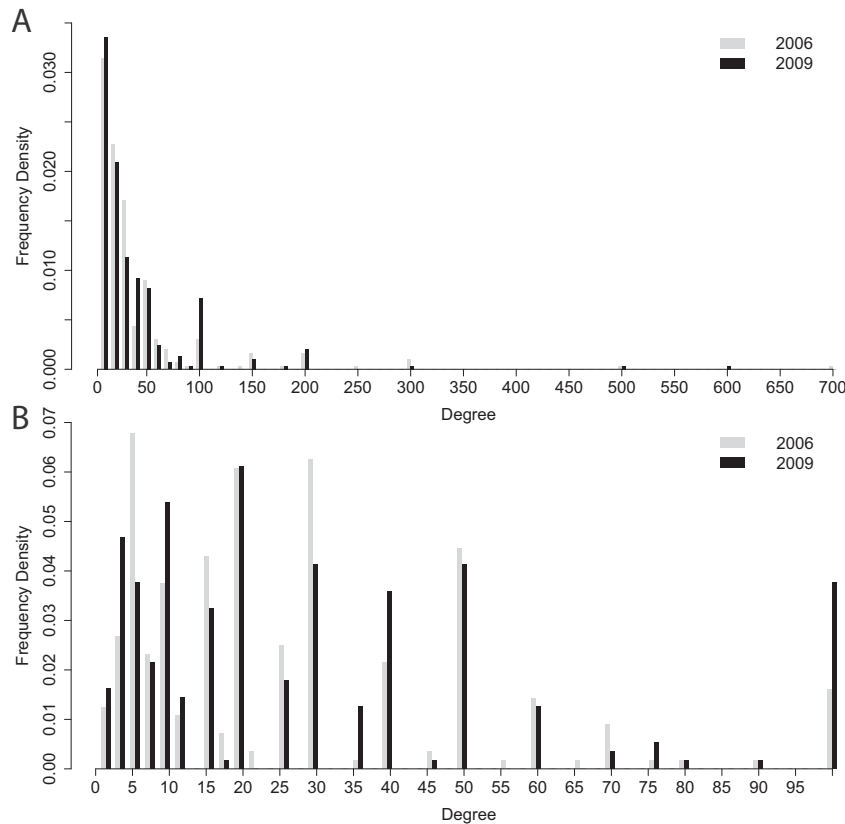
We simulate samples of size approximately 350 for each of these RDS variants, in a population of 10,000 individuals. We calculate the percentage difference between the prevalence estimates (both raw and using the Volz–Heckathorn estimator (Volz and Heckathorn, 2008)) and the actual population prevalence to determine which assumptions most impact error in RDS. We take two RDS surveys separated by two years, over a time when prevalence is increasing (from about 20% to 30%, see Fig. S4) and determine how accurately consecutive samples can identify changes in prevalence. We compare the true simulated population prevalence (prevalence in the modelled population) to the raw RDS sample prevalence and the prevalence after adjustment with the Volz–Heckathorn estimator.

## 3. Results

### 3.1. Reported contact numbers

Data describing the reported degrees in the Bristol surveys illustrate a pronounced preference of individuals to report their numbers of contacts to the nearest 10, 20, 30... and 100, 200, 300 (Fig. 1). However, it is likely that the true distribution of the numbers of relevant contacts has nearly as many 21s as 20s, nearly as many 31s and 30s and so on. The reported degree distribution is highly unlikely.

Since we only have the reported degrees, we cannot know what the true distribution is nor the details of how individuals modify this information. However, if we can generate degrees with a smooth distribution and show that, by applying a given rounding



**Fig. 1.** Degree estimates from two RDS surveys of PWIDs in Bristol, UK [20, 31], the question was 'In the last 30 days how many people who inject drugs have you spoken to who you know by name and who also know you by name?'. (A) The full distributions. (B) Just degrees less than 100. The 2006 and 2009 distributions were very similar: the mean degree in both years was 39, with a standard deviation of 66.5 in 2006 and 60.0 in 2009.

scheme, the resulting modified distribution resembles the Bristol data, we have some justification both for the choice of original distribution and the rounding scheme in question. With this objective, in Supplementary Text S4 we define a simple measure of distance between distributions. It is not immediately obvious how close two distributions should be to be considered similar. But fortunately we have two sets of reported degrees for Bristol at different times; we can therefore use the distance between the two, which we call  $z_B$ , as a yardstick for resemblance.

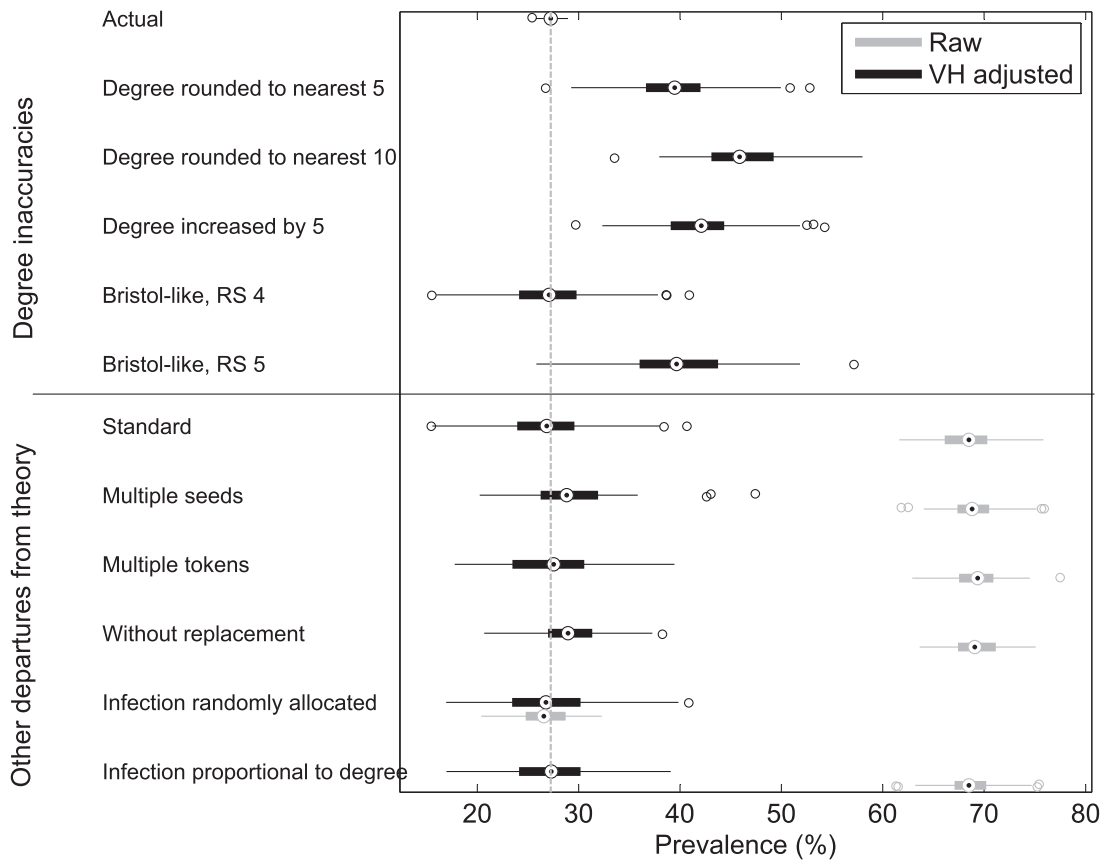
We generated long-tailed degree distributions using the power law with exponential cut-off described in Section 2, and found that the average distance to the empirical distributions was about 5.2 times  $z_B$ . We then applied each of the rounding schemes described in Section 2. Scheme 1 (rounding all degrees up by 5) and Scheme 2 (by 10), reduced the factor from 5.2 to 2.7 and 2.3, respectively. Scheme 3 (adding 5 to every degree) increased the distance somewhat to 5.5. However, the more sophisticated Scheme 4 (rounding to the nearest 10 for  $k < 100$  and to the nearest 100 for  $k > 100$ ) reduced the factor to 1.4; while Scheme 5, which is like Scheme 4 but also draws all degrees under 10 from the combined Bristol distributions, decreases this factor further still to 1.2, Table S3. In other words, these schemes produce distributions almost as close to the empirical ones as the two Bristol datasets are to each other. Note, however, that the level of interference involved in Schemes 4 and 5 should be seen as the minimum reporting error required to obtain realistic reported distributions from smooth underlying ones. If in fact it were the individuals with few contacts who nonetheless claimed to have hundreds while the highly connected reported only a small number, this would not be evident in the data. The bias introduced by inaccuracies in reported degrees which we go

on to analyse should therefore be regarded as a lower bound to the potential importance of this effect.

### 3.2. RDS simulations

Inaccuracy in reported degrees had a large effect on the reliability of estimates of prevalence and incidence (Fig. 2). The top half of Fig. 2 shows estimates of prevalence from RDS surveys where degree was mis-reported by the 5 rounding schemes. The estimates were calculated using the Volz–Heckathorn estimator. Mis-reporting degrees caused all surveys to over-estimate prevalence (compare to the 'Actual' prevalence in the whole network, top). However, if degrees were correctly reported (standard RDS) the average prevalence estimate from 100 surveys was accurate, but individual variation was large. Even with inaccurate degrees, the adjusted estimates (blue bars) were still closer to the true prevalence or incidence than the point estimate from the raw data (green bars).

Two of our degree-biasing rounding schemes were based on degrees collected in Bristol, UK. Scheme 4 adjusted only those degrees larger than 10: the prevalence estimate is comparable with the estimate using correct degrees. However, the error increased when inaccuracies were added to the lower degrees ( $1 \leq d \leq 10$ ) in Scheme 5. Those with low degree have a higher weighting in the estimator (Eq. (1)) than those with high degree; therefore mis-reporting these degrees had a larger effect on the estimate. The average prevalence for rounding Scheme 5 was 39.8% [31.1–51.4% 95% CI] compared to the actual average prevalence of 27.2% [26.1–28.4% 95% CI]. If only the high degrees were mis-reported the average estimated prevalence is closer to the actual prevalence,



**Fig. 2.** Boxplots indicating the prevalence estimates from simulated RDS surveys of 100 populations on networks with long-tailed degree distributions. The adjusted prevalence estimates (black) are always better than the raw, unadjusted data (grey), but incorrect degrees can cause significant inaccuracies in the estimated prevalence. The true prevalence in the simulated populations is indicated by the dashed line.

however separate simulations have a large variation in estimates (average 27.1% [17.0–38.6% 95% CI]). As low-degree individuals are far less likely to be infected than high-degree individuals, error in their degrees affects the denominator in the estimator without a comparable effect on the numerator (recall Eq. (1)).

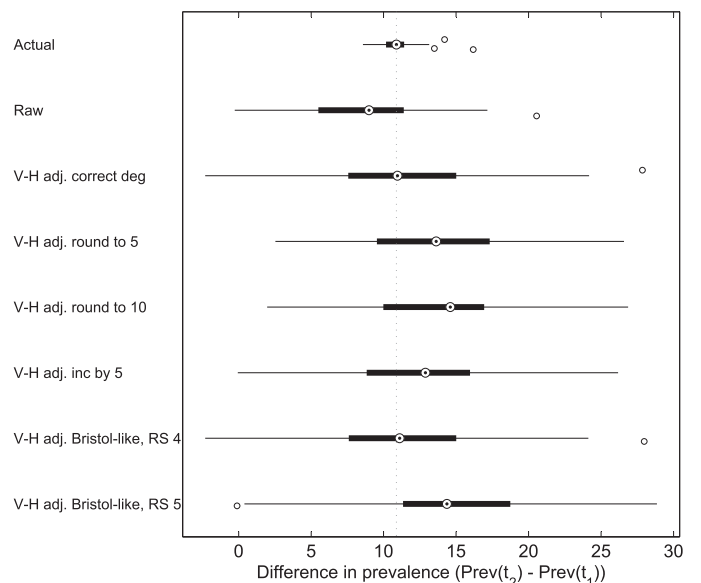
The extent of bias due to mis-reporting degrees also depends on the contact network in the population being studied. Networks with a long-tailed degree distribution have a strong correlation between degree and infection status and we showed that inaccurate degrees can lead to substantial bias in the estimation of sero-prevalence. However, in a network with a lower variance in degrees, such as a Poisson degree distribution, the correlation was much weaker and there was a correspondingly weaker effect on prevalence estimates when degrees were inaccurate (see Fig. S6).

Other differences between real RDS surveys and the idealised circumstances from which the mathematics of RDS is derived did not cause any substantial error in the estimates of incidence or prevalence in our simulations (lower half of Fig. 2). The fact that real RDS surveys use sampling without replacement, multiple original seeds rather than just one, and that individuals can recruit more than one contact into the study can be expected to lead to relatively minor differences between the theoretically unbiased estimates and estimates from data.

### 3.3. Consecutive RDS surveys

While nearly all of the simulated consecutive samples correctly identified that prevalence increased between the two times points, there were marked differences in the estimates of the sampled trend (Fig. 3 and Table S5). The true simulated prevalence on

average increased from 19% to 30% in the two year gap. The raw sample data overestimated prevalence, but identified the trend correctly. When the sample was adjusted using the Volz–Heckathorn estimator with accurate degrees the average difference across all



**Fig. 3.** Box plots illustrating the extent of the over-estimates of the increase in prevalence between two consecutive RDS surveys of the same population. Consecutive samples were taken in 100 different populations, the actual prevalence increased from 19% to 30% in the 2 year gap: this difference of 11% is indicated by the dashed line across the plot. The scenarios are fully described in the text.

100 repeats was very close to the actual increase (3rd boxplot from the top in Fig. 3). However, the variation between individual paired samples was very large, indicating large inaccuracies in individual runs. As repeated samples are impractical in reality, this implies that conclusions from consecutive studies have a high probability of being quite inaccurate, even if degrees are correctly given. This is the case for all of the rounding methods we compared.

All scenarios with mis-reported degrees had a large variance in the estimated trend in prevalence. When degrees were rounded up to 5, to 10 or increased by 5, the prevalence and the increase in prevalence between the two surveys was over-estimated. When only degrees larger than 10 were adjusted (rounding Scheme 4) the estimator returned very similar trends to those seen when adjusting with accurate degrees, indicating that the method did not remove enough of the information in degree to alter the effect of the weightings. Rounding Scheme 5, however, which added inaccuracy to those with degrees <10, showed a large average overestimate and variation in results. This indicated that it is particularly important to obtain correct degrees for low degree individuals as even small inaccuracies can have a large impact on results. The same simulation on networks with a Poisson degree distribution (and therefore a lower variance in degrees) showed a lower average over-estimate but still a large variation in results, Fig. S7.

#### 4. Discussion

There is a clear indication in the reported degrees of the Bristol data that individuals round or bin their number of contacts to the nearest 5, 10 and 100. Indeed, these empirical distributions were part of the motivation for this work; high frequencies of degrees that were multiples of 5, 10, etc. suggest that individuals may be guessing or rounding their reported degree. We analysed the effect of rounding schemes on the degree distribution and showed that schemes which round degrees to the nearest order of magnitude result in degrees with a distribution close to that seen in the Bristol data.

It is well-known that the Volz–Heckathorn adjustment reliably recovers prevalence and incidence estimates in the presence of over-sampling of high-degree individuals, in contrast to raw RDS data. However, we have found that the necessity of weighting individuals' contributions by their reported degree can lead to significant bias if degrees are inaccurately reported. This source of bias is very likely greater than inaccuracies resulting from other variations in RDS (e.g., with- or without-replacement sampling, multiple or single recruitment).

Our results highlight the importance of obtaining correct degrees for accurate analysis of RDS surveys. This has been described previously, but the extent of the effect of inaccurate degrees, particularly on serial estimates using RDS, has not been determined (Burt and Thiede, 2012; McCreesh et al., 2012; Rudolph et al., 2013; Wejnert, 2009). We find that it is particularly important to obtain correct degrees for individuals reporting low degrees. Their contribution to the estimated prevalence is high for two reasons: (1) their lower degree results in a higher weight in Eq. (1), and (2) they are less likely to be infected, so their contribution affects the denominator of the estimate without affecting the numerator. The effect of inaccurate degrees depends on the nature of the network itself, and is more pronounced where there is a stronger association between the number of contacts and the risk of becoming infected. One practical implication of this finding is that pilot studies could help to determine whether the contact network has highly variable degrees or not. If it does, then obtaining good information about the true degree of low-degree individuals will improve the accuracy of RDS-derived estimates. If not, then the effects we have reported here will be smaller.

Critically, our simulations imply that inaccurate reporting of degrees could make it difficult to assess how HCV or HIV prevalence and incidence, or other outcomes, are changing over time. In our simulations the correct (i.e., increasing rather than decreasing) trend was typically identified, but the magnitude of the trend was inaccurate. For example, it is plausible from our results that paired RDS surveys would over- or under-estimate any change by 25%. This has important implications for studies using RDS to measure the impact of interventions on HIV or HCV prevalence and incidence in PWID populations (Degenhardt et al., 2010; Martin et al., 2013, 2011; Solomon et al., 2013). For the purposes of estimating a change, researchers should consider using raw RDS values as well as adjusted ones.

The issues we report will potentially affect studies which follow the same RDS recruited individuals over time (Rudolph et al., 2011), rather than using a repeat survey; the overall number of individuals accessed will be smaller (than if multiple samples were taken), and any problems with recruitment in the initial RDS survey will persist throughout (such as difficulty reaching equilibrium). As estimates should still be adjusted using reported degrees, inaccuracies in the degrees will cause inaccurate estimates of the trends over time. Problems will occur both if the same reported degrees are used and if new reported degrees are obtained – the potential for error in the reported degrees is high.

Though we consider only increasing prevalence, the same problems will apply to populations with decreasing prevalence and to surveys taken at different time intervals. Testing all realistic permutations is not feasible, but based on our results we expect that inaccurate degrees will introduce bias into RDS surveys, and confidence in the estimates will be low, causing uncertainty in the calculated trends from paired samples.

We note that our methods of adding inaccuracy to degrees are fairly conservative; it is likely that realistic biasing behaviour is heterogeneous across a population and may depend on factors like gender, age, behaviour, degree or disease status (Bell et al., 2007; Brewer, 2000; Marsden, 1990; Rudolph et al., 2013). For example, men usually report a far higher number of sexual partners than women, giving inconsistency in the number of sexual partnerships that could have occurred (Brown and Sinclair, 1999; Liljeros et al., 2001; Smith, 1992). Similar problems may occur among PWIDs recalling injecting partnerships. In addition, PWIDs in different countries or regions where different laws and restrictions apply may bias their answers differently. These more systematic inaccuracies will likely cause a larger error in estimates, enhanced by correlations between those factors and infection. Testing the accuracy of reported degrees would be very challenging in the “hidden populations” in which RDS is used. Because of these complications, we are uncertain as to whether developing a general method to recover accurate degrees is realistic.

Most RDS questionnaires ask a variety of questions about degree, such as how many PWIDs they know by name and have seen in the last X days, how many PWIDs have they injected with in the last Y weeks and how many of these were new partners or regular partners. A recent study asking multiple questions about degree determined that the first of these obtained the most accurate answers, though mis-reporting was common (Wejnert, 2009). However, it may be possible to alter the questions to gain a more accurate understanding of an individual's risk (Rudolph et al., 2013), or to use some combination of answers to determine an alternative weighting for use with the Volz–Heckathorn estimator (Lu, 2013). Alternatively, self completion of the contact portion of the questionnaire may improve accuracy of answers (Schroder et al., 2003). Our simulations show that it is most crucial to obtain accurate reports of degree for low-degree individuals. If questioned in detail, this group may be more likely to remember contacts more accurately and may better be able to answer questions about

contact numbers, times and type of contact than individuals with dozens of contacts.

Not surprisingly, our simulations indicated that the variation in results decreases if the sample size is increased (see Figs. S5 and S6, also shown in Goel and Salganik, 2009; Mills et al., 2012). Additionally, taking multiple surveys of the same population can improve the estimate. However, multiple surveys are generally not practical. If instead a larger survey were taken, the error in estimates could be reduced by using a bootstrapping method, as described by Salganik (2006). Researchers would estimate  $\hat{\mu}$  many times, each time using a subset of the (larger) RDS sample. The resulting distribution of estimates would be used to construct confidence intervals, for example, and ultimately  $p$ -values for any estimated change in prevalence, incidence or other estimate.

We have shown that inaccuracy in degree can reduce the accuracy of prevalence estimates from RDS surveys and decrease the ability to identify accurately the magnitude of prevalence trends in the underlying population. We recognise that RDS is an extremely useful method to quickly access hidden populations such as PWIDs, but we urge users to consider results cautiously and to make every effort to estimate degrees carefully, particularly those of low-degree individuals, and particularly when comparing surveys.

### Role of funding source

H.L.M. would like to acknowledge funding from Wellcome Trust University Award 093488/Z/10/Z. S.J. is grateful for financial support from the European Commission under the Marie Curie Intra-European Fellowship Programme PIEF-GA-2010-276454. M.H. would like to acknowledge funding from NIQUAD MRC grant G1000021 and National Institute for Health Research (NIHR)'s School for Public Health Research (SPHR). N.S.J. acknowledges support from EPSRC grant EP/I005765/1. C.C. would like to acknowledge EPSRC grant EP/I031626/1 and EP/K026003/1. The funding sources had no role in the study design or analysis.

### Contributors

H.L.M., S.J., C.C., N.S.J. designed the study. H.L.M. ran the model and statistical analysis. C.C., N.S.J., S.J. and M.H. advised on the analysis. M.H. provided the datasets. H.L.M. and S.J. analysed the datasets. H.L.M. wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

### Conflict of interest

None declared.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.drugalcdep.2014.06.015>.

### References

- Abdul-Quader, A.S., Heckathorn, D.D., McKnight, C., Bramson, H., Nemeth, C., Sabin, K., Gallagher, K., Des Jarlais, D.C., 2006. Effectiveness of respondent-driven sampling for recruiting drug users in New York City: findings from a pilot study. *J. Urban Health* 83, 459–476.
- Bell, D.C., Belli-McQueen, B., Haider, A., 2007. Partner naming and forgetting: recall of network members. *Soc. Netw.* 29, 279–299.
- Bengtsson, L., Thorson, A., 2010. Global HIV surveillance among MSM: is risk behavior seriously underestimated? *AIDS* 24, 2301–2303.
- Brewer, D.D., 2000. Forgetting in the recall-based elicitation of personal and social networks. *Soc. Netw.* 22, 29–43.
- Broadhead, R.S., Heckathorn, D.D., Altice, F.L., Van Hulst, Y., Carbone, M., Friedland, G.H., O'Connor, P.G., Selwyn, P.A., 2002. Increasing drug users? Adherence to HIV treatment: results of a peer-driven intervention feasibility study. *Soc. Sci. Med.* 55, 235–246.
- Broadhead, R.S., Heckathorn, D.D., Weakliem, D.L., Anthony, D.L., Madray, H., Mills, R.J., Hughes, J., 1998. Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. *Public Health Rep.* 113, 42–57.
- Brown, N.R., Sinclair, R.C., 1999. Estimating number of lifetime sexual partners: men and women do it differently. *J. Sex Res.* 36, 292–297.
- Burt, R.D., Thiede, H., 2012. Evaluating consistency in repeat surveys of injection drug users recruited by respondent-driven sampling in the Seattle area: results from the NHBS-IDU1 and NHBS-IDU2 surveys. *Ann. Epidemiol.* 22, 354–363.
- Degenhardt, L., Mathers, B., Vickerman, P., Rhodes, T., Latkin, C., Hickman, M., 2010. Prevention of HIV infection for people who inject drugs: why individual, structural, and combination approaches are needed. *Lancet* 376, 285–301.
- Des Jarlais, D.C., Arasteh, K., Perlis, T., Hagan, H., Abdul-Quader, A., Heckathorn, D.D., McKnight, C., Bramson, H., Nemeth, C., Torian, L.V., Friedman, S.R., 2007. Convergence of HIV seroprevalence among injecting and non-injecting drug users in New York City. *AIDS* 21, 231.
- Fenton, K.A., Johnson, A.M., McManus, S., Erens, B., 2001. Measuring sexual behaviour: methodological challenges in survey research. *Sex. Transm. Infect.* 77, 84–92.
- Fisher, R.J., 1993. Social desirability bias and the validity of indirect questioning. *J. Consumer Res.* 20, 303–315.
- Friedman, S.R., Kottiri, B.J., Neaigus, A., Curtis, R., Vermund, S.H., Des Jarlais, D.C., 2000. Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. *Am. J. Epidemiol.* 152, 913–922.
- Gile, K.J., Handcock, M.S., 2010. Respondent-driven sampling: an assessment of current methodology. *Sociol. Methodol.* 40, 285–327.
- Goel, S., Salganik, M.J., 2009. Respondent-driven sampling as Markov chain Monte Carlo. *Stat. Med.* 28, 2202–2229.
- Goel, S., Salganik, M.J., 2010. Assessing respondent-driven sampling. *Proc. Natl. Acad. Sci. USA* 107, 6743–6747.
- Heckathorn, D.D., 1997. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc. Probl.* 44, 174–199.
- Heckathorn, D.D., 2002. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc. Probl.* 49, 11–34.
- Heckathorn, D.D., 2007. Extensions of respondent-driven sampling: analyzing continuous variables and controlling for differential recruitment. *Sociol. Methodol.* 37, 151–207.
- Hickman, M., Hope, V., Coleman, B., Parry, J., Telfer, M., Twigger, J., Irish, C., Macleod, J., Annett, H., 2009. Assessing IDU prevalence and health consequences (HCV, overdose and drug-related mortality) in a primary care trust: implications for public health action. *J. Public Health* 31, 374–382.
- Hope, V., Hickman, M., Ngui, S., Jones, S., Telfer, M., Bizzarri, M., Ncube, F., Parry, J., 2011. Measuring the incidence, prevalence and genetic relatedness of hepatitis C infections among a community recruited sample of injecting drug users, using dried blood spots. *J. Viral Hepat.* 18, 262–270.
- Hope, V., Hickman, M., Parry, J., Ncube, F., 2013. Factors associated with recent symptoms of an injection site infection or injury among people who inject drugs in three English cities. *Int. J. Drug Policy* 25, 303–307.
- Hope, V., Jeannin, A., Spencer, B., Gervasoni, J., van de Laar, M., Dubois-Arber, F., 2010. Mapping HIV-related behavioural surveillance among injecting drug users in Europe, 2008. *Euro Surveill.* 16, 977–986.
- Iguchi, M.Y., Ober, A.J., Berry, S.H., Fain, T., Heckathorn, D.D., Gorbach, P.M., Heimer, R., Kozlov, A., Ouellet, L.J., Shoptaw, S., Zule, W.A., 2009. Simultaneous recruitment of drug users and men who have sex with men in the United States and Russia using respondent-driven sampling: sampling methods and implications. *J. Urban Health* 86 (Suppl. 1), 5–31.
- Johnston, L.G., Malekinejad, M., Kendall, C., Iuppa, I.M., Rutherford, G.W., 2008. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. *AIDS Behav.* 12, 131–141.
- Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Aberg, Y., 2001. The web of human sexual contacts. *Nature* 411, 907–908.
- Lu, X., 2013. Linked ego networks: improving estimate reliability and validity with respondent-driven sampling. *Soc. Netw.* 35, 669–685.
- Malekinejad, M., Johnston, L.G., Kendall, C., Kerr, L.R.F.S., Rifkin, M.R., Rutherford, G.W., 2008. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behav.* 12, 105–130.
- Marsden, P.V., 1990. Network data and measurement. *Annu. Rev. Sociol.* 16, 435–463.
- Martin, N.K., Vickerman, P., Grebely, J., Hellard, M., Hutchinson, S.J., Lima, V.D., Foster, G.R., Dillon, J.F., Goldberg, D.J., Dore, G.J., et al., 2013. HCV treatment for prevention among people who inject drugs: modeling treatment scale-up in the age of direct-acting antivirals. *Hepatology* 58, 1598–1609.
- Martin, N.K., Vickerman, P., Hickman, M., 2011. Mathematical modelling of hepatitis C treatment for injecting drug users. *J. Theor. Biol.* 274, 58–66.
- McCreech, N., Frost, S.D.W., Seeley, J., Katongole, J., Tarsh, M.N., Ndungu, R., Jichi, F., Lunel, N.L., Maher, D., Johnston, L.G., et al., 2012. Evaluation of respondent-driven sampling. *Epidemiology* 23, 138–147.

- Mills, H., Colijn, C., Vickerman, P., Leslie, D., Hope, V., Hickman, M., 2012. Respondent driven sampling and community structure in a population of injecting drug users, Bristol, UK. *Drug Alcohol Depend.* 126, 324–332.
- Newman, M.E.J., 2003. The structure and function of complex networks. *SIAM Rev.* 45, 167–256.
- Robinson, W.T., Risser, J.M.H., McGoy, S., Becker, A.B., Rehman, H., Jefferson, M., Griffin, V., Wolverton, M., Tortu, S., 2006. Recruiting injection drug users: a three-site comparison of results and experiences with respondent-driven and targeted sampling procedures. *J. Urban Health* 83, 29–38.
- Rudolph, A.E., Fuller, C.M., Latkin, C., 2013. The importance of measuring and accounting for potential biases in respondent-driven samples. *AIDS Behav.* 17, 2244–2252.
- Rudolph, A.E., Latkin, C., Crawford, N.D., Jones, K.C., Fuller, C.M., 2011. Does respondent driven sampling alter the social network composition and health-seeking behaviors of illicit drug users followed prospectively? *PLoS ONE* 6, e19615.
- Salganik, M.J., 2006. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J. Urban Health* 83, 98–112.
- Salganik, M.J., Heckathorn, D.D., 2004. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol. Methodol.* 34, 193–240.
- Schroder, K.E., Carey, M.P., Vanable, P.A., 2003. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. *Ann. Behav. Med.* 26, 104–123.
- Smith, T.W., 1992. Discrepancies between men and women in reporting number of sexual partners: a summary from four countries. *Biodemography Soc. Biol.* 39, 203–211.
- Solomon, S.S., Lucas, G.M., Celentano, D.D., Sifakis, F., Mehta, S.H., 2013. Beyond surveillance: a role for respondent-driven sampling in implementation science. *Am. J. Epidemiol.* 178, 260–267.
- Sypsa, V., Paraskevis, D., Malliori, M., Nikolopoulos, G.K., Panopoulos, A., Kantzanou, M., Katsoulidou, A., Psychogiou, M., Fotiou, A., Pharris, A., et al., 2014. Homelessness and other risk factors for HIV infection in the current outbreak among injection drug users in Athens, Greece. *Am. J. Public Health* (epub ahead of print).
- Volz, E., Heckathorn, D.D., 2008. Probability based estimation theory for respondent driven sampling. *J. Off. Stat.* 24, 79–97.
- Wejnert, C., 2009. An empirical test of respondent-driven sampling: point estimates, variance, degree measures and out of equilibrium data. *Sociol. Methodol.* 39, 73–116.